

Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

5

AS 21
H 752/75
Copy 2

A Computer-Processible Nomenclature For Gene Symbols

U.S.D. Library
MAR 14 1980
PROCESSED
CURRENT SERIAL RECORDS

Contents

Introduction	1
Information storage and retrieval	1
Gene symbol nomenclature	1
Tomato gene list	2
World collection of the genus <i>Lycopersicon</i>	4
Conclusions and future research	12
Literature cited	12

A Computer Processible Nomenclature For Gene Symbols

Allan L. Shafan,¹ Raymond L. Clark,² Willis H. Skrdla,³ and Kanti Rawal¹

Introduction

Since 1947, when the North Central Regional Plant Introduction Station was established in Ames, Iowa, the genetic resources available there have been cataloged by hand. A seed list of available tomato lines was produced by typing, proofreading, and correcting stencils and by subsequent mimeographing and assembling of copies. Distribution of these mimeographed seed lists represented the final step in disseminating information about these genetic resources.

This mode of disseminating information now is inadequate. Each year, after completing the annual accessions evaluation, the seed list required updating. This entailed integrating the new data and subsequently revising the seed list. Further, many cooperating scientists maintained files of hand- or type-written reports in various formats, and these data often were not included in the seed lists.

Manual procedures also proved inadequate in processing requests for tomatoes with specific combinations of useful characteristics. These requests were handled by visually scanning the seed list and noting those accessions with the desired character combinations. When the collection was small (less than 1,000 lines) and only two traits were considered, selecting desired accessions could be accomplished easily.

Now, however, the tomato collection contains more than 4,000 lines, and we frequently receive requests for material with several traits. With manual techniques, these requests are difficult to

handle. As a result, much useful data and many useful genetic resources are not adequately utilized.

Information Storage and Retrieval

Computer-assisted methods have proved useful in facilitating the use and exchange of information. Among the various computer tools is an information storage and retrieval (IR) system that simplifies extracting a subset from the information in a data collection. Individual IR systems have unique features, but generally, they allow the user to obtain selected information from the entire data collection. When plant genetic resources are considered, selection criteria typically are combinations of plant traits.

To process data with a computerized IR system, several preliminary steps are required: (a) The user must decide which variables (characters) are to be considered, (b) a meaningful scheme for representing values for these variables must be determined, (c) data must be prepared in a machine-readable form, and (d) the machine-readable data must be processed by the IR system. Many IR systems keep stored data in an internal format, which is called a data bank. Once prepared, the data bank is easily interrogated and updated.

Existing data on a World Collection of the Genus *Lycopersicon* raised an interesting point. Not only were identification and morphological, pathological, and horticultural observations recorded, but also data from gene analysis. For storing genetic data, gene symbols must be represented in a computer-processible nomenclature.

Gene Symbol Nomenclature

Gene symbolization is based on an established system, but considerable variation exists in published nomenclature (2, 4, 5, 6, 7). Communication among breeders and geneticists could be enhanced

¹ Systems analyst, Information Sciences and Genetic Resources, University of Colorado, Boulder 80302.

^{2,3} Plant pathologist and agronomist, Regional Plant Introduction Station, U.S. Department of Agriculture, Science and Education Administration—Agricultural Research, Iowa State University, Ames 50011.

⁴ Chief scientist, Information Sciences and Genetic Resources, University of Colorado, Boulder 80302.

greatly by using a standard system for nomenclature (7). The Tomato Genetics Cooperative (TGC) has published rules for nomenclature in tomato genetics (2, 3, 4), and we used these as the basis for the computer-processible nomenclature for gene symbols described in this paper. Table 1 contains the nomenclatorial rules set forth by the Gene List Committee of the TGC (2, 4).

Computing machines have certain limitations regarding printable characters or symbols; therefore, we made minor modifications in the TGC genetic nomenclature to produce a computer-processible nomenclature. The suggested changes in the nomenclature are listed in table 2. In summary, these changes reflect the availability of superscripts and italics. Modified notation for superscripts include two forms: "¹" and "²". The symbol "¹" has been used in this paper. The modification concerning parentheses is not a general computing limitation but is imposed by the IR system that we used.

One additional nomenclatorial rule, not explicitly discussed in the current TGC rules, concerns mimics, which makes the use of subscripts unnecessary. As shown in subsequent examples, however, cross-referencing obsolete symbols may be necessary. In such cases, we used the symbol "~" or "□" to indicate that a subscript follows. Thus, *d*₂ becomes *d*~2 or *d*□2. The two alternative forms of notation reflect machine differences.

To demonstrate the applications of computer-processible notation for gene symbols, we constructed two data banks by using the Executive Information Retrieval System (EXIR). One contains information extracted from the list of tomato genes published by the TGC (3), and the second contains information on accessions in the World Collection of the Genus *Lycopersicon* (5). Only subsets of available data have been used to limit the output volume in the examples.

The term "descriptor" is defined as a variable or attribute that is observed on a set of elements in a data bank. The values for a descriptor are referred to as "descriptor states." Each data bank will be discussed individually.

Tomato Gene List

Descriptors included in the gene list data bank reflect the format of the report of the Gene List Committee in TGC Report No. 21 (3), and they demonstrate the effectiveness of an IR system for cross-referencing. The descriptors SYMBOL and

NAME follow the rules for nomenclature and identify the mutant gene under consideration. The descriptor PRIMARY CLASS and descriptors CLASS 1 through CLASS 21 are adaptations of the classification scheme devised by the TGC to seek lines of a specific phenotype (3). The classification key is given in table 3. Each line is classified according to its salient phenotypic characters and may occur in several of the 21 phenotype classes. For example, a line with yellow and white variegation will fall in classes 2 (white), 3 (yellow), and 7 (variegated) (3). The PRIMARY CLASS is the one most significant to the mutant program, and it can be any class between 1 and 21. Valid descriptor states for the descriptors CLASS 1 through CLASS 21 are PRIM, YES, and NO, respectively, for instances in which a class is primary for the line, is affirmative but not primary, or is negative. SEED SOURCE 1 and SEED SOURCE 2 are coded descriptors that indicate where stocks with the gene can be obtained. TGC REF contains the number of the TGC report that contains more information. The descriptor LINKAGE GROUP shows the chromosome number on which the gene is located (TGC Report No. 23). Descriptors OBSOLETE SYMBOL 1 through OBSOLETE SYMBOL 4 provide cross-reference to obsolete or superseded symbols. Examples of data printouts by using these descriptors are shown in tables 4 and 5. The gene symbol lists were produced by the Report Generator Module of EXIR (EXIR, RPG), and as genetic analyses continue, the gene data bank will require updating and production of updated reports.

In the information retrieval examples that follow, only a subset (59) of the entire gene list data bank was used to keep the output volume small and simple.

The genes can be ordered in various ways, and table 6 displays the genes classified by linkage group. Genes on this list were selected by the EXIR expression NOT SYMBOL, UNKNOWN, which means all genes with a known (not unknown) gene symbol. In a display (EXIR PRINT), the notation "----" indicates that the descriptor state is unknown for the descriptor under consideration. Therefore, the last genes in this list have unknown linkage groups.

To retrieve information on all stocks in one repository, classified by primary phenotypic classes, the query shown in table 7 could be processed. The capability for retrieving currently accepted gene symbols for former symbols is illustrated in

Table 1.—Rules for nomenclature in tomato genetics (2, 4)

1. Chromosomes. The chromosomes are numbered according to their length measured in pachytene. Such numbers have already been applied by Barton (1); chromosome 1 is the longest, chromosome 12, the shortest. In addition to length, such features as position of centromere and amount and distribution of heterochromatin serve to identify each chromosome. Short arms are symbolized by "S," long ones by "L"; thus, "1S" designates the short arm of chromosome 1.

2. Linkage groups. Linkage groups bear the same numbers as their respective chromosomes. As soon as the arm location of a gene is known, the locus numbering shall be revised to reflect the information. The smaller arm of each chromosome is designated as the left arm, and the zero position is the distal or left end of the small arm.

3. Genes. Mutant genes are designated by letter symbols. The mutant name, an adjective or noun or a combination of both, refers to the main diagnostic feature of the phenotype. The initial letter of the symbol should be the same as that of the name; additional appropriate letters are added as necessary to distinguish it from other symbols already in use.

The genes *c*, *r*, *s*, and *y* do not conform to this rule, their symbols having unfortunately been derived from the normal instead of the mutant condition. Since these symbols have been in widespread use for over 40 years, they should be retained, but all others should conform to these rules.

After obtaining reasonable evidence for the existence of a new gene for which the phenotype can be distinguished reasonably well in some or all genotypic milieux, the discoverer should select an appropriate name and symbol, taking care not to use any already reported.

Special effort should be expended to find the appropriate names for new mutants. Latin names or their English derivatives are recommended for their international meaning.

The symbol should consist of the minimum letters or numbers necessary to distinguish it from all other existing symbols. Although lengthy symbols may have reference value to the discoverer, they are cumbersome for genetic notation and become a hindrance to other workers, who often make greater use of the mutants than the discoverer.

4. Alleles. Dominance or recessiveness of a mutant gene is indicated by comparison with a "standard" or "normal" type. The variety Marglobe is proposed as this normal type since it is widely grown and is typical of the general concept of normal tomato morphology.

A mutant gene dominant to the normal type is written with the initial letter of the name and symbol capitalized, while a recessive is written with all letters in lower case. The normal allele of a mutant gene is written with the symbol of the mutant gene followed by the superscript "+." Thus the normal allele of *sp* is *sp*⁺ and of *l* is *l*⁺. A dominant allele appearing later at the *sp* locus would be designated *sp*⁰. Additional alleles at the same locus are designated by appropriate letter or number superscripts; thus for the *d* locus, the following alleles are known: *d*⁺, *d*^x, *d*^r, and *d*⁺. For the first member of a numbered series of alleles, which is the preferable form of notation, the "1" shall be understood but not used. When it is clear in the text which gene is concerned, the normal allele may be designated simply by the "+" symbol.

5. Indistinguishable alleles of independent origin (supposed reoccurrences). We recommend that these not be given unique symbols. If they are, however, they shall be designated by the

existing gene symbol, which shall be enclosed in parentheses, shall consist of an Arabic numeral corresponding to the particular unique abbreviation derived from the discoverer's name, this abbreviation to be assigned by the Gene List Committee to the discoverer and used by him for designating reoccurrences of any locus. For example, *ag*^(1K) and *ag*^(2K) (formerly *ag*³ and *ag*⁴ respectively), the first and second reoccurrences of *ag* found by Knowles. If the gene symbol already has a superscript, this shall precede the appended parenthetical superscript.

6. Mimics. Mutants that have different loci and are indistinguishable phenotypically from each other are known as mimics. Members of a single mimic series are preferably designated by different names and symbols (for example, *u* for uniform, *ug* for uniform gray-green). If a numbered series is used, the same name and base symbol are followed by a hyphen and a distinguishing number of the same type level. For the first member of such a mimic series, the "1" shall be understood but not used (for example, *ms*, *ms*-2, and so forth).

7. Translocations are designated by the symbol "T." The chromosomes involved are designated by their respective numbers. To distinguish between translocations involving the same chromosomes, lowercase letters are used following the chromosome numbers, thus *T*(1-2)*a*, *T*(1-2)*b*, and so forth.

8. Inversions are designated by the symbol "In", while the chromosome in which the inversion occurs is indicated by its respective number. Lowercase letters are used to distinguish different inversions on the same chromosome, thus *In*(1)*a*, *In*(1)*b*, and so forth.

9. Deficiencies are designated by the symbol "Df" and are distinguished in the same manner as inversions (rule 8).

10. Aneuploids. Monosomics and primary trisomics are designated according to the missing or extra chromosome; thus "triplo-1" refers to the primary trisomic of chromosome 1, and "hoplo-12" refers to the monosomic for chromosome 12.

For aneuploids possessing whole arm interchanges, the chromosomes are symbolized according to their component arms, a connective dot representing the centromere; thus "1S·12L" specifies an interchange between the short arm of chromosome 1 and the long arm of chromosome 12. According to this system, *2n*+10L·10L would be an example of a secondary trisomic, *2n*+5S·7L an example of a tertiary trisomic, *2n*+8L an example of a tetrasomic, and *2n*-3S·3L+3S·3L an example of a compensating trisomic in which a normal chromosome is compensated by isochromosomes of its short arm and long arm.

11. To distinguish between gene symbols and symbols of the chromosome aberrations, the former are italicized.

12. Since chromosomes of all investigated species of *Lycopersicon* appear to be almost completely homologous, we suggest that the same symbolization apply to the entire genus. We also propose that the complete gene symbols not be duplicated among the species unless the genes in question are known to be identical, and that the key letters of the symbols not be duplicated unless the genes are alleles or mimics. Species alleles can be designated by a superscript to indicate the species, for example, *a*^h for a hypothetical allele in *L. hirsutum*.

13. Much confusion will be avoided by clearing names and symbols with the chairman of either the Gene List Committee or Coordinating Committee of the TGC at the earliest opportunity. Such action will settle matters of priority and assist the investigator by assigning a reasonably permanent symbol to his mutant. Permanence cannot be guaranteed, however, because unknown

Table 1 continued on next page.

Table 1.—Rules for nomenclature in tomato genetics (2, 4)—Continued

earlier or simultaneous publication might establish priority.

14. Modifiers. Modifiers (mo or Mo) of specific genes shall be symbolized by the appropriate symbol followed by the symbol of the gene modified enclosed in parentheses. Subsequent nonallelic modifiers shall be denoted by adding a number suffix to the combined symbol. For example, the two dominant modi-

fiers of *I* are *Mo(I)* and *Mo(I)2*. When the action of the modifying gene is known to be suppression, use a suppressor symbol (*Sup*); likewise, use an enhancer symbol (*Enh*) for genes of enhancing action. In either case, the gene whose expression is modified shall be symbolized in parentheses as specified for modifiers, thus, *Enh(I)*, *Sup(I)2*.

table 8 by using the currently accepted symbol **-ah-** for the mutant formerly known as **ao**.

As a further example, we asked for a list of mutants that affect plant habit and size and either leaf necrosis or leaf form and size. The retrieved information shown in table 9 is ordered by SYMBOL, and other information includes the descriptors NAME, PRIMARY CLASS, and SEED SOURCE 1.

Table 2.—Modification to the rules for nomenclature in tomato genetics

(Refer to table 1)

1. No change
2. No change
3. No change
4. Superscripts are not available. Either the symbol "{'" or "{'" has been used to indicate that a superscript follows. The following allelic series illustrates this modification:

TGC	Computer-processible
d	d
d ^x	d x or d x
d ^{cr}	d cr or d cr
d ⁺	d + or d +

5. As noted in number 4 above, an alternate notation for superscripts has been employed. One further modification affects this rule. The symbols "<" and ">" have been substituted for left and right parentheses respectively. The following series of reoccurrences illustrates this modification:

TGC	Computer-processible
ag ^(1K)	ag <1K> or ag <1K>
ag ^(2K)	ag <2K> or ag <2K>

6. No change
7. As noted in number 5 above, an alternate notation for parentheses has been employed. Thus, T(1-2)a becomes T<1-2>a.
8. Same as number 7.
9. Same as number 7.
10. No change
11. Italics are not available. The distinction between gene symbols and symbols of the chromosomal aberrations should be derived from context.
12. For species alleles, the alternate superscript notation described in number 4 has been employed.
13. No change
14. Same as number 7.

World Collection of the Genus *Lycopersicon*

Descriptors in the *Lycopersicon* data bank that have been chosen demonstrate how various traits can be combined to selectively retrieve information on lines that constitute any desired subset of the germplasm bank. Identification descriptors include PI NUMBER and SOURCE, and 12 genes were chosen to describe plant genotypes. The genes chosen were dictated by their usefulness for illustrative purposes. The chosen loci were **B**, **c**, **I**, **Mi**, **ms-9**, **ms-17**, **pe**, **r**, **sp**, **Sw**, **Ve**, and **Wo**. The descriptor states that each locus can assume are the gene symbols for the alleles identified at that locus. Thus, for the descriptor **Wo**, the valid descriptor states are **Wo**, **Wo_m**, **Wo_v** and **+** (an abbreviated representation of **Wo₊**).

Table 3.—Key to mutant classes (3)

Class	Description
1	Anthocyanin modification: intensification, reduction, elimination
2	Chlorophyll deficiency: white or whitish
3	Chlorophyll deficiency: yellow or yellowish
4	Chlorophyll deficiency: light grey or dull grey
5	Chlorophyll deficiency: yellow-green
6	Chlorophyll deficiency: virescent (localized at growing point).
7	Irregular variegation, flecking or striping
8	Leaf necrosis
9	Hair modification: augmentation, reduction, distortion, elimination
10	Leaf form and size
11	Plant habit and size
12	Flower form and color
13	Inflorescence exclusive of 12
14	Sterility: any condition leading to partial or complete unfruitfulness
15	Fruit form and surface texture
16	Fruit color and flavor
17	Disease resistance
18	Miscellaneous characters: earliness, wilting, root mutation, and so forth
19	Seed
20	Foliage color: dark green
21	Foliage color, miscellaneous: olive, brown, blue-green

Table 4.—List of tomato mutant genes

GENE	MUTANT NAME	P	C	S	S	S	S	T	R	L	G	OBSOLETE	OBSOLETE	OBSOLETE	OBSOLETE
SYMBOL		R	L	E	O	E	I	R				SYMBOL 1	SYMBOL 2	SYMBOL 3	SYMBOL 4
		I	A	E	U	E	C	F	N	O	K	U			
		M	S	D	R	D	R	C	A	P					
		A	S	C	C	E	E	G							
		R	Y	1	2										
a	anthocyaninless	1	C	NC	4	11									
aa	anthocyanin absent	1	WK	Lat	20	2									
ac	apocarpous	15	ex		9										
acr	acroantha	4	S		17										
acu-2	accumbens-2	10	S		15										
ad	alternaria resistance	17	W	K	4										
adp	adpressa	11	S	C	12	7									
adp-2	adpressa-2	11	S		15										
adu	adusta	8	S		17										
adu-2	adusta-2	8	S		20										
adu-3	adusta-3	8	S		20										
ae	entirely anthocyaninless	1	Ro		9							a-332			
aeg	aegrota	8	S	R	9										
aeg-2	aegrota-2	8	S		20										
af	anthocyanin-free	1	K		9	5						a-325			
afe	afertilis	14	S		17										
af1	albifolium	2	R	He	12	4									
afr	anthocyaninless-fragile	1	K	Ro	12										
ag	anthocyanin gainer	1	B	R	4	10									
ag <1K>	anthocyanin gainer <1K>	1	R		17							ag 3			
ag <1S>	anthocyanin gainer <1S>	1	R		17							ag 2			
ag <2K>	anthocyanin gainer <2K>	1	R		17							ag 4			
ah	Hoffman's anthocyaninless	1	R	Ro	9								a-337		
ai	incomplete anthocyanin	1	Ro		9							a-342			
ai 2	incomplete anthocyanin 2	1	Ro		12							am	a-340		
al	anthocyanin loser	1	B	Ro	4	8						a-2			
ala	albina	2	S	La	15							alb			
ala-2	albina-2	2	S	La	15							alb-2			
ala-3	albina-3	2	S	La	15							alb-3			
ala-4	albina-4	2			17										
alb	albescens	7	T		15	12									
alu	alutacea	3	S		17										
alv	alboviridis	12	S		20							alb			
an	anantha	12	S		4							an 1	an 2	ca	
an-2	anantha-2	10	S		17										
ana	angusta	4	S	Ro	20										
ang-2	angustifolia-2	4	S		15										
ant	aurantia	4	S		20										
anu	angusta	10	S		20										
ap	apetalous	12	R	NC	4	11									

Table 5.—List of mutant tomato genes

[illegible]

Descriptors for vine and foliage characters include VINE SIZE, GROWTH, HABIT, and LEAFLET SIZE. Fruit character descriptors include CONCENTRIC CRACKING, RADIAL CRACKING, MINIMUM FRUIT SIZE, MAXIMUM FRUIT SIZE, SHAPE, SET, MATURITY, FLESH INTENSITY, MIN NO LOCULES, MAX NO LOCULES, and FASCIATION.

Tables 10 and 11 display data for some accessions in the World Collection. These tables were produced by the Report Generator Module of EXIR (EXIR RPG), and they show the types of report that can be produced by a computerized IR system.

Such lists often are useful, but the greatest ability of the IR system is its capacity to retrieve information only on accessions of interest. For example, if one wants the *sp* allele occurring at the *sp* locus and the normal (+) allele at the *I* locus in accessions that are early to medium maturing (1 to 3), information on the selected accessions can be displayed. Table 12 shows the descriptors MATURITY, MAX NO LOCULES, GROWTH HABIT, and PI NUMBER for these accessions. Note that the output is ordered hierarchically, first by MAX NO LOCULES and then by GROWTH HABIT, and PI NUMBER, respectively.

Table 6.—ID: Sample exir query on the tomato gene list data bank

PRINT: LINKAGE GROUP, (SYMBOL, NAME, PRIMARY CLASS, OBSOLETE 1)
FOR GENES WITH NOT SYMBOL, UNKNOWN*

NO. OF ITEMS IN QUERY RESPONSE = 59

NO. OF ITEMS IN THE DATA BANK = 59

PERCENTAGE OF RESPONSE/TOTAL DATA BANK = 100.00

1	au <1S> au 3	aurea <1S> aurea 3	3 au 3 au-2
2	aa aw	anthocyanin absent without anthocyanin	1 --- 1 aba
3	aut	aureata	--- ---
4	af1	albifolium	2 ---
5	af	anthocyanin-free	1 a~325
7	adp atv	adpressa atroviolacium	11 --- 1 ---
8	al	anthocyanin loser	1 a~2
9	ah	Hoffman's anthocyaninless	1 ao
10	ag	anthocyanin gainer	1 ---
11	a ap	anthocyaninless apetalous	1 --- 12 ---
12	alb aud	albescent auroid	7 --- 3 ---
---	ac acr acu-2 ad adp-2 adu adu-2 adu-3 ae aeg aeg-2 afe afn ag <1K>	apocarpous acroxantha accumbens-2 alternaria resistance adpressa-2 adusta adusta-2 adusta-3 entirely anthocyaninless aegrota aegrota-2 afertilis anthocyaninless-fragile anthocyanin gainer <1K>	15 --- 4 --- 10 --- 17 --- 11 --- 8 --- 8 --- 8 --- 1 a~332 8 --- 8 --- 14 --- 1 --- 1 ag 3

Table 7.—ID: Sample exir query on the tomato gene list data bank

PRINT: PRIMARY CLASS, SYMBOL
FOR GENES WITH SEED SOURCE 1,R OR SEED SOURCE 2,R*

NO. OF ITEMS IN QUERY RESPONSE = 13
NO. OF ITEMS IN THE DATA BANK = 59
PERCENTAGE OF RESPONSE/TOTAL DATA BANK = 22.03

1
ag
ag|<1K>
ag|<1S>
ag|<2K>
ah
atv
aw|<1K>
2
af1
3
aud
8
aeg
12
ap
16
at

aut

Table 8.—ID: Sample exir query on the tomato gene list data bank

PRINT: SYMBOL FOR THE GENE WITH OBSOLETE 1,ao*

NO. OF ITEMS IN QUERY RESPONSE = 1
NO. OF ITEMS IN THE DATA BANK = 59
PERCENTAGE OF RESPONSE/TOTAL DATA BANK = 1.69

ah

Table 9.—ID: Sample exir query on the tomato gene list data bank

PRINT: (SYMBOL, NAME, PRIMARY CLASS, SEED SOURCE 1)
FOR GENES WITH CLASS 11,PRIM OR YES
AND (CLASS 8,PRIM OR YES OR CLASS 10,PRIM OR YES)*

NO. OF ITEMS IN QUERY RESPONSE = 13
NO. OF ITEMS IN THE DATA BANK = 59
PERCENTAGE OF RESPONSE/TOTAL DATA BANK = 22.03

acr	acroxantha	4	S
acu-2	accumbens-2	10	S
adp	addressa	11	S
adp-2	addressa-2	11	S
adu	adusta	8	S
adu-2	adusta-2	8	S
adu-3	adusta-3	8	S
afe	afertilis	14	S
ana	angusta	10	S
ang-2	angustifolia-2	4	S
ant	aurantia	4	S
apl	applanata	10	S
atn	attenuata	5	S

Table 10.—World collection of the genus *Lycopersicon*

P. I.	NUMBER	SOURCE	B	C	I	Mi	ms-9	ms-17	pe	r
	254654	KERR-1-12	+	+	+	+	+	+	+	+
	254655	KERR-1-M	+	+	+	+	+	+	+	+
	254656	BURDICK 220-19	+	C	I	+	+	+	+	+
	254657	BUTLER 12-5	+	+	+	+	+	+	+	+
	254658	P.A. YOUNG G1734C-10	+	+	+	+	+	+	+	+
	254659	HOFFMAN GLOBE A	+	+	+	+	+	+	+	+
	265955	GILBERT HES 6360 ANAHU	+	+	+	Mi	+	+	+	+
	265956	GILBERT KOLEA C	+	+	+	Mi	+	+	+	+
	265957	GILBERT HES 6321 <6147>	+	+	+	Mi	+	+	+	+
	273444	M. HARDIN 'BUSHY'	+	+	+	+	+	+	+	+
	279565	CARD RED	B	+	+	+	+	+	+	+
	279566	P.A. YOUNG SV616C	+	+	+	+	+	+	+	+
	279567	PEARL HARBOR, C.O.CLAYBERG2172	+	+	+	+	+	+	+	+
	279568	P.A. YOUNG SR1234E <L.PIMP.>	+	+	+	+	+	+	+	+
	289763	P.A. YOUNG PINKDEAL STEP 329	+	C	+	+	+	+	+	R
	293592	CLARK'S EARLY DOHLER NO.7CANAD	+	+	+	+	+	+	+	+
	293593	GENEVA NO 6 CANADA	+	+	+	+	+	+	+	+
	293594	EARLIEST OF ALL CANADA	+	+	+	+	+	+	+	+
	293595	NON-ACID CANADA	+	+	+	+	+	+	+	+
	293596	RED CHIEF CANADA	+	+	+	+	+	+	+	+

Table 11.—World collection of the genus *Lycopersicon*

P. I.	NUMBER	sp	Sw	Ve	Wo	V S I I N Z E E	G R D W T H	H A B I T	L S E I A Z F E L E T	C C O R N A C C E K N T G	R C A R D A I C A K L N G	MIN FRUIT SIZE	MAX FRUIT SIZE	S H A P E	M A T U R T Y	F I L N T S N H S T Y	M N L I U O C B U E L R E S	M N L A U O X M C B U E L R E S	F A S C I A T
	254654	+	+	+	+	2		ERECT	3	2		2.00	2.00	1	3	3	MANY	MANY	4
	254655	+	+	+	+	1		ERECT	3		2	1.50	1.50	2	1	5	2	2	1
	254656	+	+	+	+	4		INTER	3		2	1.75	1.75	1	5	4	MANY	MANY	4
	254657	sp	+	+	+	2		INTER	3			1.50	1.50	2	5	3	2	3	1
	254658	+	+	+	+	4		INTER	5	3	3	3.00	3.00	1	3	4	MANY	MANY	3
	254659	+	+	+	+	4		INTER	4	2	2	3.00	3.00	2	4	4	5	5	1
	265955	+	+	+	+	4		INTER	2	2	2	2.00	2.50	2	3	5	MANY	MANY	1
	265956	+	+	+	+	2	IND	INTER	3	1	5	2.00	2.50	2	4	3	MANY	MANY	1
	265957	+	+	+	+	3	DET	INTER	3		2	2.00	2.50	2	3	5	MANY	MANY	1
	273444	sp	+	+	+	1	DET	INTER	1	3	2	2.00	2.50	2	5	2	5	5	1
	279565	+	+	+	+	4	IND	INTER	4	3	3	2.50	3.00	2	4	4	4	MANY	1
	279566	+	+	+	+	1	IND	INTER	1	1	1	1.50	2.00	5	3	1	3	3	1
	279567	+	Sw	+	+	3	IND	INTER	1	3	2	1.50	2.50	2	5	1	2	MANY	2
	279568	+	+	+	+	4	IND	INTER	2	1	1	.50	.75	2	3	4	3		
	289763	Sp	+	+	+	2	IND	INTER	3	2	2	1.75	2.50	2	3	4	4	MANY	1
	293592	+	+	Ve	+	3	IND	INTER	1	3	3	2.00	2.50	2	3	3	4	MANY	2
	293593	+	+	Ve	+	1	IND	INTER	1	1	2	1.50	2.00	2	1	1	1	MANY	1
	293594	+	+	Ve	+	3	IND	INTER	3	4	3	1.25	1.25	2	4	3	4	MANY	1
	293595	+	+	Ve	+	2	IND	INTER	2	3	4	1.00	2.00	2	3	3	1	MANY	1
	293596	+	+	Ve	+	1	DET	INTER	1	1	1	1.00	1.50	2	3	3	1	MANY	1

Another example where the search was for accessions with medium to large vine size (3 to 5) and any allele other than the normal type at the *Mi* locus is shown in table 13. The user in this case chose to order the output by PI NUMBER.

Table 14 displays a query for fruit characters and genotypic information on accessions with the *r* allele at the *r* locus and any allele other than the normal type at either the *I* or *Sw* locus. The output is ordered by MINIMUM FRUIT SIZE, MAXIMUM FRUIT SIZE, SHAPE, SET, and FASCIATION, respectively.

Table 12.—ID: Sample exir query on a subset of the world Lycopersicon collection data bank

```

PRINT: MATURITY,(MAX NO LOCULES,GROWTH,HABIT,PI NUMBER)
      FOR ACCESSIONS WITH sp,sp AND I,+ AND MATURITY, FROM 1 TO 3*

NO. OF ITEMS IN QUERY RESPONSE =    7
NO. OF ITEMS IN THE DATA BANK =   83
PERCENTAGE OF RESPONSE/TOTAL DATA BANK =  8.43

1
  3    --- ERECT 205046
2
  5    DET INTER 273444
  MANY --- INTER 193421
3
  3    --- INTER 224577
  3    --- INTER 254657
  5    --- ERECT 224572
  MANY --- INTER 193411

```

Table 13.—ID: Sample exir query on a subset of the world Lycopersicon collection data bank

```

PRINT: (PI NUMBER,SOURCE,Mi,VINE SIZE )
      FOR ACCESSIONS WITH VINE SIZE, FROM 3 TO 5 AND NOT Mi,++

NO. OF ITEMS IN QUERY RESPONSE =    2
NO. OF ITEMS IN THE DATA BANK =   83
PERCENTAGE OF RESPONSE/TOTAL DATA BANK =  2.41

265955 GILBERT HES 6360 ANAHU           Mi  4
265957 GILBERT HES 6321 <6147>          Mi  3

```

Table 14.—ID: Sample exir query on a subset of the world Lycopersicon collection data bank

```

PRINT: (MINIMUM FRUIT SIZE, MAXIMUM FRUIT SIZE,SHAPE,SET,FASCIATION,I,Sw,PI NUMBER )
      FOR ACCESSIONS WITH r,r AND (NOT I,+ OR NOT Sw,+)*

NO. OF ITEMS IN QUERY RESPONSE =    5
NO. OF ITEMS IN THE DATA BANK =   83
PERCENTAGE OF RESPONSE/TOTAL DATA BANK =  6.02

1.75 1.75 1    1    2    I    +    224578
2.00 2.00 1    3    1    I    +    212429
2.00 2.00 1    4    2    I    +    193412
2.00 5.00 1    3    2    I    +    229809
3.25 3.25 1    3    4    I    +    205039

```

Conclusions and Future Research

The computer-processible nomenclature described in this paper can serve the needs of the community of scientists involved in research in tomato genetics. The nomenclatorial rules of the Tomato Genetics Cooperative were used as the basis for the computer-processible nomenclature, and the symbols chosen proved very effective in a computerized information storage and retrieval system. Display of information as well as selective retrieval and updating of information was greatly facilitated.

The gene symbol nomenclature shown for tomatoes is applicable to many crops as a vehicle for effective computer-assisted information management.

We discovered several inconsistencies and ambiguities during work with the nomenclature. Further development of the nomenclature will seek to resolve these inconsistencies. Future research will include investigation into the nomenclatorial systems being used in other crops such as oats and maize. The end product from this research will be a consistent system of nomenclature for gene symbols that is computer-processible, minimizes ambiguity, and is widely acceptable to the entire genetic research community.

Literature Cited

- (1) Barton, D. W. 1950. Pochytene morphology of the tomato chromosome complement. *American Journal of Botany* 37: 639-643.
- (2) Clayberg, C. D., L. Butler, C. M. Rick, E. A. Kerr, and R. W. Robinson. 1970. Rules for nomenclature in tomato genetics. *Tomato Genetics Cooperative Report* No. 20:3-5.
- (3) ———. 1971. Condensed list of genes listed in the past *Tomato Genetics Cooperative Reports*. *Tomato Genetics Cooperative Report* No. 21:2-10.
- (4) ———. 1973. Additional rule for nomenclature. *Tomato Genetics Cooperative Report* No. 23:3.
- (5) Simons, M. D., F. J. Zillinsky, and N. F. Jensen. 1966. A Standardized system of nomenclature for genes governing characters of oats. *U.S. Department of Agriculture, ARS* 34-85.
- (6) Skrdlo, W. H., L. J. Alexander, G. Oakes, and A. F. Dodge. 1968. Horticultural characters and reaction to two diseases of the world collection of the genus *Lycopersicon*. *Ohio Agricultural Experiment Station Bulletin* 1009. 110 p.
- (7) Tanaka, Y. (committee chairman) and others. 1957. Report of the international committee on genetic symbols and nomenclature. In *Union International des Sciences Biologiques, Comptes Rendus, Section B, Colloquio*, No. 30.

U.S. DEPARTMENT OF AGRICULTURE
SCIENCE AND EDUCATION ADMINISTRATION
NORTH CENTRAL REGION
PLANT INTRODUCTION STATION
IOWA STATE UNIVERSITY
AMES, IOWA 50011

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE \$300



POSTAGE AND FEES PAID
U.S. DEPARTMENT OF
AGRICULTURE
AGR 101